

PARTE 1

(contestar cada parte por separado)

Ejercicio 1 (1,00 pts.)

A continuación se definen dos modelos lineales 2-dimensionales:

- *Modelo 1:* $\theta_1 = \{b = 1, w_1 = 5, w_2 = 1\}$.
- *Modelo 2:* $\theta_2 = \{b = 1, w_1 = 5, w_2 = 0\}$.

Dado el siguiente conjunto de datos, compuesto por 3 patrones y dos características (x_1, x_2) :

$x_{i,1}$	$x_{i,2}$	y_i
1	0	6
0	1	2
1	1	7

- Calcular la salida de ambos modelos sobre los 3 patrones.
- Calcular el Error Cuadrático Medio (Mean Squared Error, MSE) para cada uno de los dos modelos. ¿Qué modelo es mejor según esta medida?
- ¿Cuál de los dos modelos parece corresponderse a un modelo Lasso? ¿Por qué?

Ejercicio 2 (1,50 pts.)

Dada una Máquina de Vectores Soporte (Support Vector Machine, SVM) de clasificación, definida por el siguiente problema de optimización primal, y su problema dual equivalente:

$$\begin{aligned} \min_{\substack{\mathbf{w} \in \mathbb{R}^d \\ b \in \mathbb{R} \\ \xi \in \mathbb{R}^N}} \left\{ \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^N \xi_i \right\} &\iff \min_{\alpha \in \mathbb{R}^N} \left\{ \frac{1}{2} \alpha^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \alpha - \alpha^T \mathbf{1} \right\} \\ \text{s.t. } \begin{cases} y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \\ \xi_i \geq 0, \\ 1 \leq i \leq N, \end{cases} &\quad \text{s.t. } \begin{cases} \alpha^T \mathbf{y} = 0, \\ 0 \leq \alpha \leq C. \end{cases} \end{aligned}$$

- Explicar el significado del parámetro C . ¿Valores más grandes de C producen mayor o menor tendencia al sobreajuste?
- Explicar en qué consiste en este caso el truco del núcleo (*kernel trick*)¹.
- El problema dual de las SVMs se suele resolver con el algoritmo de Sequential Minimal Optimization (SMO). Explicar cuántos coeficientes duales se actualizan en cada iteración de SMO, y por qué.

Ejercicio 3 (1,00 pts.)

Considerando las Redes Neuronales Profundas (Deep Neural Networks, DNNs) estudiadas durante el curso:

- ¿Qué tipo de función de activación se suele usar en las capas ocultas? ¿Cuáles son sus ventajas?
- Indicar de forma razonada qué arquitectura especializada se usaría para abordar los siguientes problemas:
 - Distinguir el dígito que aparece manuscrito en una imagen.
 - Generar imágenes artificiales que simulen ser imágenes reales.
 - Reconstruir imágenes eliminando el ruido.
 - Predicir la evolución de una serie temporal.
- Se quiere entrenar una DNN para que obtenga la máxima puntuación posible en un videojuego. ¿Qué tipo de paradigma de Aprendizaje Automático se puede usar y por qué?

¹No es necesario introducir formalismos, basta con una explicación intuitiva.

PARTE 2

(contestar cada parte por separado)

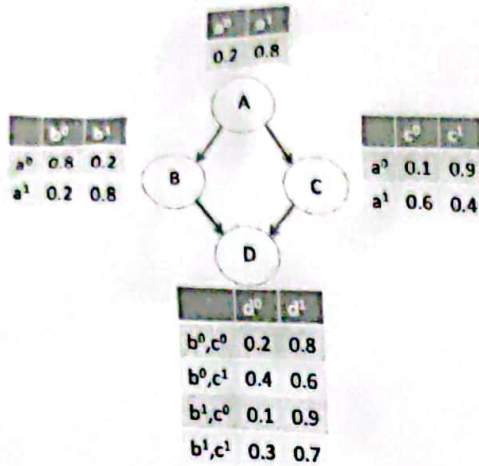


Figura 1

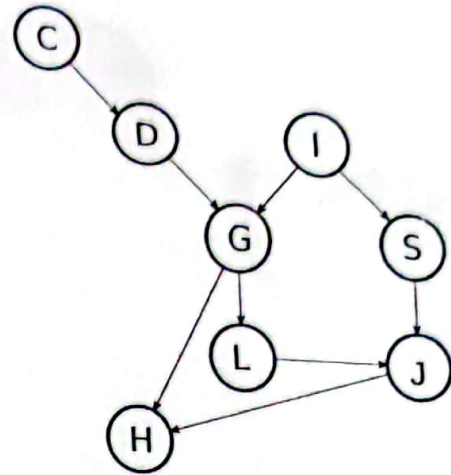


Figura 2

Ejercicio 4 (1,50 ptos.)

Se debe calcular $P(B|c = c^1)$ usando el algoritmo de eliminación de variables para la red bayesiana de la Figura 1. Para ello realizar los siguientes pasos:

1. Escribir la distribución conjunta de probabilidad $P(A, B, C, D)$ de la red.
2. Escribir $P(B|c = c^1)$ a partir de la fórmula anterior (sin cálculos numéricos).
3. A partir de la fórmula anterior elimina primero la variable D .
4. Finalmente, calcula los valores numéricos finales para $P(B|c = c^1)$ mediante la eliminación de la variable A .

Ejercicio 5 (0,50 ptos.)

Analiza bajo qué condiciones S influye a C en la red bayesiana de la Figura 2.

Ejercicio 6 (1,50 ptos.)

Responde a las siguientes cuestiones relacionadas con conjuntos de clasificadores:

1. Describe brevemente el algoritmo general de Gradient Boosting.
2. Describe brevemente el algoritmo de Random Forest.
3. Comenta las ventajas y desventajas de los algoritmos GB y RF.

PARTE 3

(contestar cada parte por separado)

Responde brevemente a las siguientes preguntas. Se valora sobre todo demostrar que se entienden los conceptos principales, no es necesario entrar en descripciones muy detalladas. Uno o dos párrafos deberían ser suficientes.

Ejercicio 7 (1,00 ptos.)

Estimación de densidades.

1. Explica el concepto de máxima verosimilitud. ¿Cuál es la estimación de parámetros mediante máxima verosimilitud de una distribución gaussiana?
2. Explica el concepto de los Gaussian Mixture Models. ¿En qué se diferencian del apartado anterior? ¿Qué problema pretenden resolver y qué algoritmo se usa para obtener la máxima verosimilitud en los GMMs?
3. Da dos ejemplos de estimación de densidades no paramétrica, y describe los conceptos de *kernel function* y *bandwidth*.

Ejercicio 8 (0,50 ptos.)

Reducción de dimensionalidad.

1. Indica dos aplicaciones de la reducción de dimensionalidad.
2. Describe brevemente las técnicas de PCA y LLE, y explica para qué tipos de datos son más apropiadas cada una.

Ejercicio 9 (1,00 ptos.)

Clustering.

1. Describe brevemente el algoritmo de Kernel k -Means.
2. ¿En qué casos es más apropiado que k -Means o GMM Clustering?

Ejercicio 10 (0,50 ptos.)

Reglas de asociación.

1. Define los conceptos de soporte y confianza de una regla de asociación. ¿Qué reglas deben seleccionarse?
2. ¿Cuál es el principal problema de eficiencia de los algoritmos de minería de reglas de asociación?